

Summary

Summary

Methodological News contains articles on a range of topics across different areas within the Methodology Division in the ABS. This issue contains 3 articles on Response Propensity Modelling with Machine Learning, A Subsampling Approach to Non-Response Follow Up and Unsupervised Machine Learning for Data Editing and Correction.

Response Propensity Modelling with Machine Learning

RESPONSE PROPENSITY MODELLING WITH MACHINE LEARNING

The difficulties in achieving and maintaining response rates are leading National Statistical Offices (NSOs) to focus on potential respondents and how to direct data collection efforts most efficiently to maximise the quality of outputs. This area of research is known as adaptive and responsive design and complements efforts to improve the level of response via improving the provider experience. ABS is currently seeking to identify a line of investigation regarding the use of machine learning for response propensity modelling.

For household surveys and before standard workload enumeration, we want to predict the propensities on how a geographic area will respond to different collection protocols. The propensities will inform decisions on protocols to use and the number and location of interviewers needed. Finally, after standard workload enumeration, we want to predict the likelihood of getting responses from those that have not yet responded and use this to assign follow-up protocols.

For business surveys we want to apply response propensity modelling before the Intensive Follow Up (IFU) stage to identify units that will respond without reminder letters or calls. This is a cost saving initiative, referred to as a "Gold Provider" strategy in previous ABS work.

During the IFU, we want to identify units least likely to respond regardless of how many additional reminders we make. These units would be de-prioritised in the IFU, allowing effort to concentrate on units that are more likely to deliver a response.

It is proposed that the Random Forests Machine Learning method be investigated. This method appears to be well suited to our problem because it deals with small sample sizes relative to the number and type of predictors as well as behaviour that involves factor interactions not known by the researcher at the time the model is estimated. The aggregation across trees is expected to generate more stable estimates compared to those generated from any single tree.

For more information, please contact Carl Mackin Methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

A Subsampling Approach to Non-Response Follow Up

A SUBSAMPLING APPROACH TO NON-RESPONSE FOLLOW UP

The ABS recognises a pressing need to reduce overall survey costs. Field costs are a very large component of overall survey costs and are driven up by the requirement to deliver high response rates. The effort required to reach response rate targets has been increasing over time. Current ABS research aims to provide a strategy for cost reduction in data collection that manages the impacts on estimate quality. In this context we consider quality in terms of sampling error (variance) and non-response error (non-response bias).

Structural cost savings can be realised, *inter alia*, with a focus on the organisation of intensive follow up (IFU)

efforts. The specific method under consideration is sub-sampling for IFU. After initial contact and a request to complete a survey are made to each unit in the selected sample, any non-responding units are currently followed up in an effort to elicit response. The idea behind sub-sampling for IFU is to instead focus the IFU effort (or a fraction of the existing IFU effort) to only a fraction of the non-responding units.

The sub-sampling approach is supported by estimation methods that make use of the distinction between units that respond to the initial request and those that respond through follow-up, or, for categorical variables, the constancy in the ratio of propensities between the initial and IFU respondents. We have the choice of some units receiving IFU with certainty (e.g. influential businesses, or units first joining a panel) while the selection probability of other units can be optimised with respect to criteria of interest (e.g. creating a sample with more even response propensities).

For more information, contact Bruce Fraser Methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

Unsupervised Machine Learning for Data Editing and Correction

UNSUPERVISED MACHINE LEARNING FOR DATA EDITING AND CORRECTION

Many administrative datasets and survey datasets have missing data and errors. There may be little subject matter knowledge available to indicate what checks need to be done to identify errors, and specifying a set of such edit rules is laborious. It is also not always clear how to impute for missing item values in a way that is consistent with reported items.

Supervised learning is already being explored in the ABS for situations where training data are available from historical editing by humans. In this situation the goal is to establish a model using the pre-edited data to predict an outcome such as "reported value is incorrect". Predictions from such a model can be used to guide future editing, but supervised machine learning can replicate biases that were inherent in the manual editing processes.

For this study, the direction proposed is to use machine learning or automated modelling techniques to explore features of a dataset, for example by fitting a model for the joint density of the reported items on a dataset. Such a model would give low probability to units with item relationships that appear infrequently in the dataset. Methods would be preferred in which the unusual relationships can be reported to a human expert who can then determine whether the situation signifies an error. This can then be incorporated in edit rules, in a revised model that can highlight items for correction and in imputing a new item value.

For more information, contact Philip Bell Methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

How to Contact Us and Email Subscriber List

HOW TO CONTACT US AND EMAIL SUBSCRIBER LIST

Methodological News features articles and developments in relation to methodology work done within the ABS Methodology Division. By its nature, the work of the division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

If you would like to be added to or removed from our electronic mailing list, please contact:

Joey Srinkapaibulaya
Methodology Division
Australian Bureau of Statistics
Locked Bag No. 10
Belconnen ACT 2617

Email: methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

About this Release

Methodological News is a quarterly information bulletin from the Methodology Division. The newsletter features articles and developments in relation to work done with the division.

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence